FINAL REPORT

FOUNDATIONS OF OBJECT DETECTION AND RECOGNITION

Brown University / PI Prof. Ulf Grenander / ARPA Grant No. MDA972-93-1-0012

August 20, 1998

Start Date: 1 July, 1993
End Date: 31 August, 1997
Funding Level: $1,068,554 Total 4-Year Funding From All Sources
($227,090 - 1 JUL 93 - 30 SEP 93; $277,893 - 1 OCT 93 - 30 SEP 94; $186,192 - 1 OCT 94 - 30 SEPT 95; $255,444 - 1 OCT 95 - 30 SEP 96; $121,934 - 1 OCT 96 - 31 AUG 97)

Principal Participants on the Project: Ulf Grenander (Brown University), Donald Geman (University of Massachusetts and Brown), Stuart Geman (Brown), Basilis Gidas (Brown), Donald McClure (Brown), Chris Raphael (Brown), E. Bienenstock (Brown and University of Paris), Anuj Srivastava (Brown and Florida State University).

OBJECTIVES:

1) Develop mathematical foundations for a unified approach to object (in particular target) detection and recognition.

2) Accommodate multi-sensor data and scenarios with large numbers of objects.

3) Accommodate rigid and non-rigid transformations, variations in lighting conditions, contrast, noise, blur, and clutter.

4) Overcome the massive computational hurdles inherent in the general problem of object detection and recognition.

NOTABLE ACCOMPLISHMENTS.

1. The starting point when *building a mathematical foundation for ATR* has been the assertion that in order to be able to see and understand scenes it is necessary to have some prior knowledge about the scene ensemble that is expected to be encountered. We construct such priors using pattern theoretic ideas and try to catch the essential characteristics of the scene ensemble through the introduction of a configuration space $\mathcal{C}$. This leads to prior

probability measures Π on $\mathcal{C}$. It should be noted, however, that there will be several such measures, each one corresponding to the knowledge we happen to have in a particular situation, say parametrized through a parameter vector $\theta \in \Theta$. The coordinates of the parameter space Θ can express for example meteorological conditions like temperature, lighting conditions, such as the position of the Sun, or tactical conditions, such as the type and number of vehicles expected in the scene when such knowledge, more or less accurate, has been made available through other means. In this way the probability measure will be conditioned into some Π_θ on $\mathcal{C}$.

In this approach to ATR an instrumental idea is the role of prior conditions made explicit through parameters θ. One could express this by saying that these parameters represent a *statistical map* of the potential scene ensemble. Typically the map will give only an incomplete description of the scene, perhaps that it is of type 'desert' without specifying the location of individual sand dunes, or that it can contain vehicles of a certain type without specifying their location and orientation, nor their number $n \geq 0$.

The potential OOI's (Objects Of Interest) are represented in more detail, say through CAD models - templates operated on by low dimensional transformation groups, generically denoted by S. The group could be for example the Euclidean group in the plane, $S3DSE(2)$, for totally rigid objects, augmented if necessary by a low number of dimensions if rigidity is not total, for example that of a rotating turret. Or, for FLIR sensors, the thermal profile is represented by a low dimensional multiplicative group depending upon a θ that expresses temperature conditions and recent object activity.

In dynamic situations, for example searching for aircraft in air space, the probabilities on the transformation groups represent what is known about the dynamics of the targets: mass, moments of inertia, limitations on thrust and torque, etc., and relate them to the development of trajectories through the equations of Newtonian mechanics. We have explored this possibility and constructed priors also allowing several targets in the scene.

This is all put together by using the compositional aspect of pattern theory, through which we combine OOI's with background, and the transformational aspect, through which we modify the resulting scenes by applications of transformations from S.

The prior knowledge has now been represented by a prior probability measure Π_θ which is then conditioned by the information acquired by the sensors, a mathematical structure, the deformed image $I^{\mathcal{D}}$, which is typically an array, not necessarily rectangular. A cross array of radars, for example, would have an output consisting of complex scalars arranged in a cross like configuration. To formalize this we often write $I^{\mathcal{D}} = 3Dn(Tsc)$ where the sensor transformation T takes the true (but unknown) configuration $sc \in \mathcal{C}$ into an array Tsc, and $n(I)$ means a noise array depending upon the image array I. In simple situations the noise can be additive Gaussian, in quantum limited situations it can be Poisson, and so on.

We have not tried to contribute to the mathematics of sensor technology since this is outside our domain of expertise. Instead we have relied upon the literature to choose $T, n(\cdot), \ldots$.

This done we get the posterior probability measure $P_\theta(dc|I^{\mathcal{D}})$ on C conditioned by the observed image(s) through a straight forward application of Bayes' theorem; this posterior measure contains all the available information.

To exploit the information we have built inference engines that synthesize the posterior measure. The engine solves the equations of a jump-diffusion process recursively and the solution has been used for target recognition and detection but could also be applied to get optimal prediction of the future behavior of the target(s). This set up allows for multiple sensors as well as for multiple targets,

2. The methodology described above has been implemented for the following scenarios:

a) One or several rigid objects - tanks, APC... - observed with FLIR using CAD models for the OOI and simulated background.

b) A flying object observed with a combination of visible light camera with high resolution radar, synthesized noise.

c) A rigid object observed with a visible light camera with stereographic projection.

3. Within this framework we have derived metrics for ATR, in particular Hilbert-Schmidt lower bounds, both for detection/recognition error probabilities and for estimation errors in the Euclidean group $S = 3DSE(2)$. Since this group has curved geometry the usual linear-quadratic metrics are not suitable. We have argued that non-convex cost functions must be used and that ambiguities in inferences, that result from the lack of convexity, must be explicitly allowed in any realistic evaluation of performance in such cases.

The analytically derived lower bounds have been compared to numerical results obtained by Monte Carlo simulation in scenario c); the results were approximately the same.

4. The computational feasibility of these inference engines has been explored and saccadic search algorithms have been designed to speed them up. This part of our work is still in a preliminary form.

5. We believe that an important component in our approach is still missing: the pattern-theoretic representation of clutter. As mentioned above we need a mathematical description of the whole scene, and this includes clutter, in order to make optimal inferences. For this reason we have begun to study clutter systematically. Earlier it was difficult to get access to real data with clutter but that has become possible via some image data bases. We have used in particular the MSTAR data base. We have started with two clutter types:

A) Forest type clutter of clustering trees

and

B) Clutter where the dominating features are roads in a landscape.

In both cases some analytical results have been obtained but it is too early to claim

success as far as realism is concerned. This work is continuing.

6. A formal framework has been developed for object modelling and image interpretation based upon ideas from the cognitive sciences. Collaboration is ongoing with neurophysiologists to test specific predictions about patterns of activity in multi-unit recordings.

Related to this is the development of a theory of computational anatomy for use in medical imaging, in particular using magnetic resonance cameras. This is being done in collaboration with radiologists, psychiatrists and neuroscientists at Washington University and Iowa University. Some concrete results were obtained for the early diagnosis of schizophrenia based on shape changes in the hippocampus.

Both of these research activities are in the form of technology transfer. They do not deal directly with ATR but are based on mathematical techniques that we have constructed for ATR. Also, some of these ideas were employed for the detection of mines in shallow water.

7. The problem of reconstructing surfaces from SAR data has been studied in the context of a particular SAR application, namely the reconstruction of data collected by the Magellan probe of Venus. It is believed that this mathematical analysis should be relevant to other uses of SAR.

8. To disseminate our findings to other researchers we have of course used the usual method of publishing: technical reports, papers in professional journals, and talks at conference proceedings. In addition we have authored two CD-ROMS. One entitled "Automated Target Recognition; A Bayesian Approach" by A. Srivastava and U. Grenander contains a fairly non-mathematical presentation of our ATR work. The other, called "Evolving Anatomies" by U. Grenander and L. Matejic discusses our approach to Computational Anatomy. Both CDs have been distributed widely but copies are still available. We plan to author further CDs based on our research in the future.

Meetings have been organized to present our results. One was a workshop on ATR and was held at Brown University in 1996. Another dealt with Computational Anatomy was organized at Washington University in 1996.

CONCLUSIONS. *The task of developing a mathematical foundation for a unified approach to object recognition has been completed to some extent. We believe that this has led to a better understanding of what is really needed in ATR theory and, partially, to how this can be realized. We now have the beginning of an ATR theory and hope that it will be exploited by interested members of the ATR community.*

BIBLIOGRAPHY.

The following is an incomplete list of papers developed under this grant or directly

influenced by our work.

1. E. Bienenstock and René Doursat: A Shape-Recognition Model Using Dynamical Links, *Network: Computation in Neural Systems,* **5** 241–258 (1994).

2. M. Cooper, Ulf Grenander, Michael Miller, Anuj Srivastava: Accommodating Geometric and Thermodynamic Variability For Forward-Looking Infrared Sensors, Proc. SPIE, Vol. 3070, Algorithms for Synthetic Aperture Radar IV, E. Zelnic, Ed., pp. 162-72, 1997.

3. D. Geman and B. Jedynak: An Active Testing Model for Tracking Roads from Satellite Images, *IEEE Trans. Pattern Anal. Mach. Intell.*, 18,1-14, 1996.

4. D. Geman and Y. Amit: Shape Quantization and Recognition with Randomized Trees,*Neural Computation*, to appear.

5. D. Geman, Y. Amit, and K. Wilder: Joint induction of shape features and tree classifiers,*IEEE Trans. Pattern Anal. Mach. Intell., to appear*

6. D. Geman and B. Jedynak: An Active Testing Model for Tracking Roads from Satellite Images, *IEEE Trans. Pattern Anal. Mach. Intell.*, 18, 1-14, 1996.

7. D. Geman and Y. Amit: Shape Quantization and Recognition with Randomized Trees, *Neural Computation*, to appear.

8. D. Geman, Y. Amit, and K. Wilder: Joint induction of shape features and tree classifiers, *Neural = Computation*, to appear.

9. D. Geman and B. Jedynak: Shape recognition and twenty questions, *Proc., RFIA '94*, Vol. 1, 21-37, 1994.

10. D. Geman, Y. Amit, and B. Jedynak: Tree-based Focusing and Face Detection, Technical Report, University of Chicago; to appear in NATO ASI Series, Springer Verlag, 1997.

11. S. Geman, D. Potter, and Z. Chi, Composition Systems. Technical Report, Division of Applied Mathematics, Brown University, 1998.

12. S. Geman, K. Manbeck, and D.E. McClure, A comprehensive statistical model for single-photon emission tomography. In: *Markov Random Fields: Theory and Applications.* Eds. R. Chellappa and A. Jain. Academic Press, Boston, 1993, 93-130.

13. B. Gidas, Metropolis-type Monte Carlo Simulation Algorithm and Simulated Annealing: Topics in Contemporary Probability and its Applications, CRC Press Series in Stochastics and Probability, 1995, pp. 159-232, ed.: J. Laurie Snell.

14. B. Gidas and A. Almeida: A Variational Method for Estimating the Parameters of MRF from Complete or Incomplete Data, Annals of Applied Probability 3 (1993) 103-136.

15. B. Gidas and A. Zelic: Object Recognition via Hierarchical/Syntactic Models, Proceedings of the 13th International Conference on Digital Signal Processing, Santorini, Greece, 1997, Vol. II, pp.836-839.

16. U. Grenander, General Pattern Theory, Oxford University Press, 1993.

17. U. Grenander: Elements of Pattern Theory, Johns Hopkins University Press, 1996.

18. U. Grenander:The Wald Lectures: Pattern Theory, Inference in Complex Systems, Inst. Math. Stat. Montreal, 1995.

19. U. Grenander: Strategies for Seeing, incomplete Tech. Rep., Brown University, 1998.

20. U. Grenander, A. Srivastava, and M.I. Miller, Multiple Target Direction of Arrival Tracking, IEEE Trans. Signal Processing, 1995.

21. U. Grenander, A. Srivastava, and M.I. Miller: Conditional mean estimation via jump-diffusion processes, IEEE Trans. Signal Processing, 1995.

22. U. Grenander, A. Srivastava, and M. I. Miller: Asymptotic Performance Analysis of Bayesian Object Recognition, submitted to IEEE Transactions on Information Theory, Feb, 1998.

23. U. Grenander, M. I. Miller and A. Srivastava: Hilbert-Schmidt Lower Bounds for Estimators on Matrix Lie Groups for ATR Accepted for publication in IEEE Transactions on PAMI. Feb, 1998.

24. H. Künsch, S. Geman, and A. Kehagias: Hidden Markov Random Fields. *Annals of Applied Probability*, 5, 1995, 577-602.

25. C. Raphael and S. Geman: A grammatical approach to mine detection. In: Detection and Remediation Technologies for Mines and Minelike Targets II, SPIE Vol. 3079, Proceedings of SPIE. A.C. Dubey and R.L. Barnard, Editors, 1997, pp316-332.

26. A. Srivastava,, M. I. Miller, and U. Grenander: Multiple Target Direction of Arrival Tracking, IEEE Transactions on Signal Processing, Vol. 43, No. 5, pp. 1282-85, May 1995

27. A. Srivastava, U. Grenander, G. Jensen, M. I. Miller: Jump-Diffusion Markov Processes on Matrix Lie Groups, submitted to Advances in Applied Probability

28. A. Srivastava, M. I. Miller, and U. Grenander: Lie Group Parameterization for Dynamics Based Prior in ATR, Sixth Digital Signal Processing Workshop, November 1995

29. A. Srivastava, M. I. Miller and U. Grenander: Ergodic Algorithms on Special Euclidean Groups for ATR, book chapter in Mathematical Techniques in Networks and systems, publisher Birkhauser. Nov, 1996.

30. A. Srivastava, M. I. Miller and U. Grenander: Inferences via Jump-Diffusion Processes for Representations on Lie Groups Submitted to Advances in Applied Probability, Aug 1996.